Masahide Sasaki[1,2], Alberto Carlini[1,2] and Richard Jozsa[3]
[1] *Communications Research Laboratory, Ministry of Posts and Telecommunications, Koganei, Tokyo 184-8795, Japan.*
[2] *CREST, Japan Science and Technology*
E-mail:psasaki@crl.go.jp

[3] *Department of Computer Science, University of Bristol, Woodland Road, Bristol BS8 1UB, England.*

We consider the quantum analogue of the pattern matching problem, which consists of classifying a given unknown system according to certain predefined pattern classes. We address the problem of quantum template matching in which each pattern class $\mathcal{C}_i$ is represented by a known quantum state $\hat{g}_i$ called a template state, and our task is to find a template which optimally matches a given unknown quantum state $\hat{f}$. We set up a precise formulation of this problem in terms of the optimal strategy for an associated quantum Bayesian inference problem. We then investigate various examples of quantum template matching for qubit systems, considering the effect of allowing a finite number of copies of the input state $\hat{f}$. We compare quantum optimal matching strategies and semiclassical strategies and demonstrate an entanglement assisted enhancement of performance in the general quantum optimal strategy.

PACS numbers:03.67.-a, 03.65.Bz, 89.70.+c

I. INTRODUCTION

Let us consider the following *pattern matching* problem. We have at our disposal a database of recorded persons' pictures, organized into different classes according to certain defined features. Now we are given a further person's picture and we want to determine to which class the person belongs. We scan the database by comparing the defined features of the given sample with those of the classes. If the patterns of the sample have a good enough matching with those of a certain class, then we can say that the person is *recognized*, that is, that the person belongs to that pattern class. We would like to consider a similar problem in the quantum mechanical context.

There are various kinds of approaches to this problem [1,2]. Given a pattern we wish to classify it relative to a predefined set of pattern classes in such a way that we maximize some suitable measure of matching. This could be formulated, e.g., in terms of vector representations, that is, a given image is discretized on a mesh and the contents in each pixel are approximated by the value of some predetermined intensity levels, say $f(i,j)$ for the (i,j)-th pixel of a two dimensional mesh. Thus a pattern is represented by a vector $\vec{f} = (f(1,1), f(1,2), ...)^T$. When an input sample $\vec{f}_0$ is given, an intrinsic feature is first extracted from it by removing noise and adjusting its size. The resulting data, say $\vec{f}$, called the *feature vector*, should be less noisy, less redundant and more invariant under commonly encountered variations and distortions. We want then to classify this $\vec{f}$ into a pattern class from a set of classes $\{\mathcal{C}_i\}$, where each class $\mathcal{C}_i$ contains some similar patterns. Classification is usually made by evaluating a *discriminant function* $D_i(\vec{f})$ associated with the pattern class $\mathcal{C}_i$, and is such that if the input sample is actually in the class j, the value $D_j(\vec{f})$ must be the largest. One way to make the problem more tractable is to represent each class $\mathcal{C}_i$ by a typical pattern called a *template vector* $\vec{g}_i$ and to deal only with this vector as the representative of the class $\mathcal{C}_i$. This is the template matching problem.

Now consider a similar problem in the quantum mechanical context. We are given a feature state $|f\rangle$, which is usually unknown, and a set of classes $\{\mathcal{C}_i\}$ and their associated template states $|g_i\rangle$, known *a priori*. The problem of *quantum template matching* is to classify the state $|f\rangle$ according to the set $\{|g_i\rangle\}$, that is, to pick up the template which best matches with $|f\rangle$. This is similar in some respects to quantum state discrimination and quantum state estimation. In quantum state discrimination, a discrete set of states $\{\hat{\rho}_i\}$ and their *a priori* probabilities $\{p_i\}$ are given. The task is to decide which state is received. In quantum state estimation, one is to reconstruct a given unknown state $\hat{\rho}$ by estimating certain parameters. In both scenarios, one usually minimizes a certain Bayes cost such as a decision or estimation error by using prior knowledge about the states. In quantum template matching, although we deal with an unknown input state generally specified by continuous parameters, the purpose is not to estimate the input state or to discriminate among the input states themselves, but to assign the best matched template state from amongst given candidates. In this sense quantum template matching involves aspects of both state estimation and discrimination: the unknown input states are generally parameterized by continuous parameters which we wish to characterize only up to some approximation given by the "closest" template state. Indeed direct state estimation or discrimination would provide a strategy for template matching (by comparing the classical information of the estimated state with the classical information of the template identities) but this is generally not optimal – we should attempt to best match a template without necessarily obtaining any further more detailed information about the identity of the input state itself.

In this paper we will set up a precise formulation of this problem (in section II) in terms of a suitable intuitive matching criterion. The template matching problem will then appear as a problem of determining the optimal strategy for an associated quantum Bayesian inference problem [5,6]. We will then consider some examples of template matching for qubit systems (in sections III and IV) in particular, considering the effect of allowing a finite number of identical copies of the input state $|f\rangle$ (of course in a classical context this makes no difference). We will compare the optimal strategy (allowing full use of entanglement across the space of all copies) with two semiclassical strategies:
(a) applying only separate measurements on each copy and processing the outcomes to decide the best matching;
(b) applying the optimal state estimation strategy [12–14] (using a collective measurement on the product state of all copies) and then classically comparing the identity of the reconstructed state with that of the template states.
We will see that the optimal (fully entangled) strategy is more efficient than either of these. Finally in section V we will summarize our results and describe some interesting further possible generalizations of the concept of quantum template matching.

II. BAYESIAN FORMULATION OF TEMPLATE MATCHING

The Bayesian formulation is based on an *a priori* knowledge about the inputs: the input feature state $\hat{f}$ is unknown but it is assumed that we know the *a priori* probability distribution $P(\hat{f})$ for possible inputs. Each template state $\hat{g}_i$ representing the class $\mathcal{C}_i$ is assumed to be completely known. In order to classify $\hat{f}$ into a class $\mathcal{C}_j$, we need to introduce a *score* $S(\mathcal{C}_j|\hat{f})$ which provides a matching criterion. One reasonable choice for the score in template matching is that derived from the similarity criterion

$$S(\mathcal{C}_j|\hat{f}) \equiv \left(\mathrm{Tr}\left(\sqrt{\sqrt{\hat{f}}\hat{g}_j\sqrt{\hat{f}}} \right) \right)^2, \tag{1}$$

which is just the fidelity between the input state $\hat{f}$ and the template state $\hat{g}_i$. If the template states are pure states $|g_j\rangle$ then this is just the standard overlap $\langle g_j|\,\hat{f}\,|g_j\rangle$. Under the similarity criterion, we are to choose the template for which the state overlap with $\hat{f}$ is largest. The matching strategy is represented by a probability operator measure (POM) $\{\hat{\Pi}_j\}$:

$$\hat{\Pi}_j = \hat{\Pi}_j^\dagger \geq 0, \quad \sum_j \hat{\Pi}_j = \hat{I}. \tag{2}$$

This should be designed by using the *a priori* knowledge on the input and the template states, and the conditional scores $S(\mathcal{C}_j|\hat{f})$. The performance of a matching strategy is measured by the *average score* defined as

$$\bar{S} \equiv \sum_j \int df\, S(\mathcal{C}_j|\hat{f}) P(\mathcal{C}_j|\hat{f}) P(\hat{f}), \tag{3}$$

where $P(\mathcal{C}_j|\hat{f}) \equiv \mathrm{Tr}\,(\hat{\Pi}_j\hat{f})$ is the conditional probability that we have the j-th outcome given the state $\hat{f}$. The best strategy is the one that maximizes this average score. If we introduce the score operators

$$\hat{W}_j \equiv \int df\, S(\mathcal{C}_j|\hat{f}) P(\hat{f}) \hat{f}, \tag{4}$$

then Eq. (3) can be rewritten as

$$\bar{S} = \sum_j \mathrm{Tr}\,(\hat{W}_j\hat{\Pi}_j). \tag{5}$$

Thus the problem is to find the optimal POM $\{\hat{\Pi}_j\}$ that maximizes $\bar{S}$ given the set of score operators $\{\hat{W}_j\}$. This is a standard quantum Bayesian optimization problem and necessary and sufficient conditions for optimality are well known [3–5]:

$$\begin{aligned} &\text{(i)}\;\; \hat{\Gamma} \equiv \sum_j \hat{W}_j \hat{\Pi}_j \text{ is hermitian,} \\ &\text{(ii)}\;\; \hat{\Gamma} - \hat{W}_j \quad \geq 0 \quad \forall j. \end{aligned} \tag{6}$$

Note that since $S(\mathcal{C}_j|\hat{f})$ is always non-negative, we have $\hat{W}_j \geq 0$ as operators. Hence our optimal template matching problem (of optimizing eq. (5)) reduces to a standard quantum state discrimination problem – of distinguishing the mixed states $\hat{W}_j / \operatorname{Tr} \hat{W}_j$ (the normalized score operators) taken with prior probabilities $p_j = \operatorname{Tr} \hat{W}_j / \sum_j \operatorname{Tr} \hat{W}_j$. In general the optimal strategy is unknown but we will consider examples exhibiting symmetry in which optimal strategies can be given.

III. BINARY TEMPLATE MATCHING OF A TWO STATE SYSTEM

We begin with the simplest case of quantum template matching in which there are only two classes, and each class is described by a template state which is a known pure qubit state. Furthermore the input states $|f\rangle$ will be restricted to depend on only a single real feature parameter. By taking an appropriate qubit basis $\{|{\uparrow}\rangle, |{\downarrow}\rangle\}$, the binary template states can be represented in terms of real components

$$|g_0\rangle = \cos\frac{\theta}{2}|{\uparrow}\rangle + \sin\frac{\theta}{2}|{\downarrow}\rangle, \tag{7a}$$

$$|g_1\rangle = \sin\frac{\theta}{2}|{\uparrow}\rangle + \cos\frac{\theta}{2}|{\downarrow}\rangle, \tag{7b}$$

with a single real parameter θ specifying the nonorthogonality between the templates. As for the input state $|f\rangle$, we will assume that its input distribution is the uniform probability density over the great circle on the Bloch sphere defined by the two template states. Thus we can write

$$|f(\phi)\rangle = \cos\frac{\phi}{2}|{\uparrow}\rangle + \sin\frac{\phi}{2}|{\downarrow}\rangle, \tag{8}$$

where the *a priori* density of ϕ is uniform, $P(f) = P(\phi) = (2\pi)^{-1}$. We are now to decide which template is closest to the given $|f(\phi)\rangle$ in the sense of the highest state overlap. We suppose further that we are given N identical copies of the input feature state $|F(\phi)\rangle = |f(\phi)\rangle^{\otimes N}$ and the average score can be written as

$$\bar{S}(N) = \sum_{j=0}^{1} \frac{1}{2\pi} \int_0^{2\pi} d\phi \operatorname{Tr}\left(\hat{\Pi}_j \hat{F}(\phi)\right) |\langle f(\phi)|g_j\rangle|^2, \tag{9}$$

where $\hat{F}(\phi) \equiv |F(\phi)\rangle\langle F(\phi)|$. Note that our score $S(C_j|f)$ is still just $|\langle f(\phi)|g_j\rangle|^2$, the overlap for a single copy (i.e. we are establishing a relation between the input pattern $|f(\phi)\rangle$ and the templates $|g_j\rangle$) but our POM Π_j operates on the full space of N copies. The full input system is described on the $N+1$ dimensional *totally symmetric bosonic subspace* of $\mathcal{H}^{\otimes N}$, $\mathcal{H}_B$ [13,15], as

$$|F(\phi)\rangle \equiv |f(\phi)\rangle^{\otimes N} = \sum_{k=0}^{N} \sqrt{\begin{pmatrix} N \\ k \end{pmatrix}} \left(\cos\frac{\phi}{2}\right)^{N-k} \left(\sin\frac{\phi}{2}\right)^{k} |k\rangle, \tag{10}$$

where $\{|k\rangle\}$ is the occupation number basis of the $\downarrow$-component. For example, in the case of $N=3$, the basis state $|2\rangle$ reads

$$|2\rangle \equiv \begin{pmatrix} 3 \\ 2 \end{pmatrix}^{-\frac{1}{2}} \left(|{\uparrow\downarrow\downarrow}\rangle + |{\downarrow\uparrow\downarrow}\rangle + |{\downarrow\downarrow\uparrow}\rangle\right). \tag{11}$$

Our score operators are then given by

$$\hat{W}_j \equiv \frac{1}{2\pi}\int_0^{2\pi} d\phi \hat{F}(\phi) |\langle f(\phi)|g_j\rangle|^2; \quad j=0,1, \tag{12}$$

where each $\hat{W}_j$ has support on the $N+1$-dimensional subspace $\mathcal{H}_B$ and Eq. (9) can be rewritten as

$$\bar{S}(N) = \sum_{j=0}^{1} \mathrm{Tr}\,(\hat{W}_j \hat{\Pi}_j). \tag{13}$$

Without loss of generality the matching strategy $\{\hat{\Pi}_0, \hat{\Pi}_1\}$ is constructed on $\mathcal{H}_B$. In the occupation number basis representation (using eqs. (10) and (12)), the score operators are explicitly given as

$$\langle k|\,\hat{W}_j\,|l\rangle = \sqrt{\begin{pmatrix} N \\ k \end{pmatrix}\begin{pmatrix} N \\ l \end{pmatrix}}\frac{(k+l-1)!!(2N-k-l-1)!!}{(2N+2)!!}\left[(N-k-l)\cos\theta_j + (N+1)\right], \tag{14}$$

when $k+l$ is even, and

$$\langle k|\,\hat{W}_j\,|l\rangle = \sqrt{\begin{pmatrix} N \\ k \end{pmatrix}\begin{pmatrix} N \\ l \end{pmatrix}}\frac{(k+l)!!(2N-k-l)!!}{(2N+2)!!}\sin\theta_j, \tag{15}$$

when $k+l$ is odd, where $\theta_0 \equiv \theta$ and $\theta_1 \equiv \pi - \theta$.

A. Optimal template matching

Now we consider the optimal strategy that satisfies the conditions of Eq. (6). In the present case of binary classification, the analysis is rather straightforward, as we are to maximize the following quantity

$$\bar{S}(N) = \mathrm{Tr}\,(\hat{W}_0\hat{\Pi}_0) + \mathrm{Tr}\,(\hat{W}_1\hat{\Pi}_1) \tag{16}$$

$$= \mathrm{Tr}\,(\hat{W}_1) + \mathrm{Tr}\left[(\hat{W}_0 - \hat{W}_1)\hat{\Pi}_0\right], \tag{17}$$

where the resolution of the identity $\hat{\Pi}_0 + \hat{\Pi}_1 = \hat{I}$ was used in the second equality. Since $\mathrm{Tr}\,(\hat{W}_1) = 1/2$, $\hat{\Pi}_0$ should be taken to maximize $\mathrm{Tr}\left[(\hat{W}_0 - \hat{W}_1)\hat{\Pi}_0\right]$, that is, it should be the projection onto the subspace corresponding to the positive eigenvalues of the operator $\hat{W}_0 - \hat{W}_1$. From Eqs. (14) and (15) we have that

$$\langle k|\,(\hat{W}_0 - \hat{W}_1)\,|l\rangle = \sqrt{\begin{pmatrix} N \\ k \end{pmatrix}\begin{pmatrix} N \\ l \end{pmatrix}}\frac{(k+l-1)!!(2N-k-l-1)!!}{(2N)!!}\frac{(N-k-l)}{(N+1)}\cos\theta, \tag{18}$$

when $k+l$ is even, and $\langle k|\,(\hat{W}_0 - \hat{W}_1)\,|l\rangle = 0$ otherwise. Although we have not succeeded in deriving an explicit analytic expression for the eigenvalues λ_k of $\hat{W}_0 - \hat{W}_1$, we introduce the diagonalizing operator $\hat{P}$ such that

$$\hat{P}(\hat{W}_0 - \hat{W}_1)\hat{P}^\dagger = \left(\sum_{k=0}^{N} \lambda_k\,|k\rangle\,\langle k|\right)\cos\theta, \tag{19}$$

where $\lambda_0 > \lambda_1 > .... > \lambda_N$. Since $\hat{W}_0 - \hat{W}_1$ is antisymmetric in the antidiagonal (i.e. $(\hat{W}_0 - \hat{W}_1)_{kl} = -(\hat{W}_0 - \hat{W}_1)_{N-k,N-l}$) the eigenvalues match up in $\pm$ pairs: $\lambda_N = -\lambda_0$, $\lambda_{N-1} = -\lambda_1$, and so on (and when N is even, $\lambda_{N/2} = 0$). The optimal strategy can then be constructed from the pair of projection operators $\hat{\Pi}_0, \hat{\Pi}_1$ onto the subspaces of nonnegative and negative eigenvalues respectively, which can be written as:

$$\hat{\Pi}_0 = \hat{P}^\dagger \hat{\Pi}_0^{\mathrm{MV}} \hat{P}, \quad \hat{\Pi}_0^{\mathrm{MV}} \equiv |0\rangle\,\langle 0| + \cdots + \left|\lfloor\frac{N}{2}\rfloor\right\rangle\left\langle\lfloor\frac{N}{2}\rfloor\right|, \tag{20a}$$

$$\hat{\Pi}_1 = \hat{P}^\dagger \hat{\Pi}_1^{\mathrm{MV}} \hat{P}, \quad \hat{\Pi}_1^{\mathrm{MV}} \equiv \left|\lfloor\frac{N}{2}\rfloor + 1\right\rangle\left\langle\lfloor\frac{N}{2}\rfloor + 1\right| + \cdots + |N\rangle\,\langle N|, \tag{20b}$$

where $\lfloor N/2 \rfloor$ is the integer part of $N/2$. The maximum average score can be finally written as

$$\bar{S}_{\rm OPT}(N) = \frac{1}{2} + \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \lambda_k \cos\theta. \tag{21}$$

The expressions (20) for the POM also provide an intuitively appealing interpretation of the matching strategy, which consists of two steps. The first step is the unitary operation $\hat{P}$ which is applied to the N-product input state $|F(\phi)\rangle$. The second step is the measurement of the transformed state $\hat{P}\hat{F}(\phi)\hat{P}^{\dagger}$ by the POM $\{\hat{\Pi}_0^{\rm MV}, \hat{\Pi}_1^{\rm MV}\}$. This corresponds to a separate measurement in the $\{|\uparrow\rangle, |\downarrow\rangle\}$ basis on each input copy space, followed by majority voting on the outcomes. In other words, the transformation $\hat{P}$ prepares the optimal entangled state for the final measurement, and $\hat{\Pi}_0^{\rm MV}$ ($\hat{\Pi}_1^{\rm MV}$) provides the projection onto the $\uparrow$-majority (the $\downarrow$-majority) bosonic subspace. Note that for the case of a single copy (i.e. $N = 1$) $\hat{W}_0 - \hat{W}_1$ is diagonal in the $\{|\uparrow\rangle, |\downarrow\rangle\}$ basis so that $P = I$ and the measurement in this basis is the optimal strategy.

In a semiclassical strategy where the separable measurement $\{\hat{\Pi}_0^{\rm MV}, \hat{\Pi}_1^{\rm MV}\}$ (corresponding to the optimal measurement on each separate copy) is directly applied without using the transformation $\hat{P}$, the attained average score becomes, instead

$$\bar{S}_{\rm MV}(N) = \frac{1}{2} + \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \begin{pmatrix} N \\ k \end{pmatrix} \frac{(2k-1)!!(2N-2k-1)!!}{(2N)!!} \frac{(N-2k)}{(N+1)} \cos\theta. \tag{22}$$

$\bar{S}_{\rm OPT}(N)$ and $\bar{S}_{\rm MV}(N)$ are compared numerically in Fig. 1 for the case of orthogonal templates ($\theta = 0$). The effect of $\hat{P}$ can be seen to reduce the required number of sample copies to attain a prescribed level of the average score (the curve denoted by "+" corresponds to the strategy consisting of quantum state estimation and classical matching, which will be explained in the next subsection).

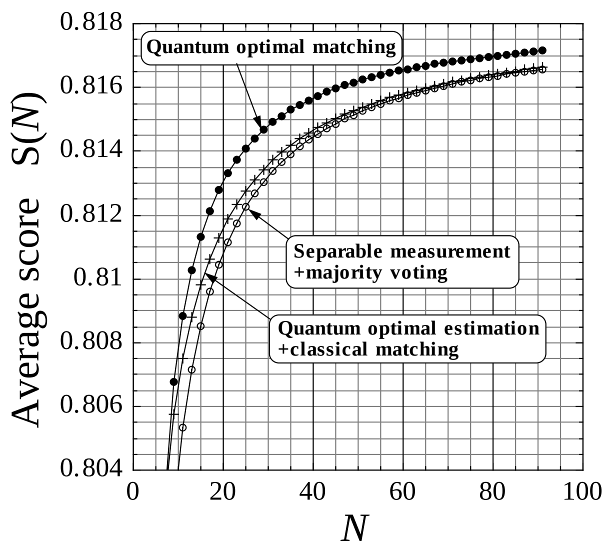

FIG. 1. The average score in the binary template matching as a function of the available number of copies of the input. Three strategies are compared. The black circles represent $\bar{S}_{\rm OPT}(N)$ for the optimal strategy, while the white circles represent $\bar{S}_{\rm MV}(N)$ for the strategy of separable measurement + majority voting. The plus correspond to $\bar{S}_{\rm EST}(N, N+1, \frac{\pi}{M})$ for the strategy of the optimal state estimation + classical matching (section III B).

Let us illustrate the optimal matching strategy in the case where we use three sample copies. The operator to be diagonalized is

$$\hat{W}_0 - \hat{W}_1 = \frac{\cos\theta}{2^6} \begin{pmatrix} 15 & 0 & \sqrt{3} & 0 \\ 0 & 3 & 0 & -\sqrt{3} \\ \sqrt{3} & 0 & -3 & 0 \\ 0 & -\sqrt{3} & 0 & -15 \end{pmatrix}. \tag{23}$$

The diagonalizing matrix is found as

$$\hat{P} = \begin{pmatrix} \cos\gamma & 0 & \sin\gamma & 0 \\ 0 & \cos\gamma & 0 & -\sin\gamma \\ -\sin\gamma & 0 & \cos\gamma & 0 \\ 0 & \sin\gamma & 0 & \cos\gamma \end{pmatrix}, \tag{24}$$

where $\cos\gamma = [(2\sqrt{21}+9)/4\sqrt{21}]^{1/2}$ and $\sin\gamma = [(2\sqrt{21}-9)/4\sqrt{21}]^{1/2}$. We then have

$$\hat{P}(\hat{W}_0 - \hat{W}_1)\hat{P}^\dagger = \frac{\cos\theta}{2^6} \begin{pmatrix} \sqrt{21}+3 & 0 & 0 & 0 \\ 0 & \sqrt{21}-3 & 0 & 0 \\ 0 & 0 & -\sqrt{21}+3 & 0 \\ 0 & 0 & 0 & -\sqrt{21}-3 \end{pmatrix}, \tag{25}$$

and

$$\bar{S}_{\mathrm{OPT}}(3) = \frac{1}{2} + \frac{\sqrt{21}}{2^4}\cos\theta. \tag{26}$$

One possible circuit structure for the optimal classifier is shown in Fig. 2. The input state $|F(\phi)\rangle$ is first transformed by $\hat{P}$, and is then processed interactively with two ancillary qubits via two controlled-NOT and two controlled-controlled-NOT gates. These steps implement $\{\hat{\Pi}_0^{\mathrm{MV}}, \hat{\Pi}_1^{\mathrm{MV}}\}$ as a measurement on a single qubit. By measuring the second ancillary qubit $|\sigma\rangle_X$ in the basis $\{|\uparrow\rangle, |\downarrow\rangle\}$, we can decide with the maximum average score that the best matched template is $|g_0\rangle$ (respectively $|g_1\rangle$) when the output is $|\uparrow\rangle$ (respectively $|\downarrow\rangle$).

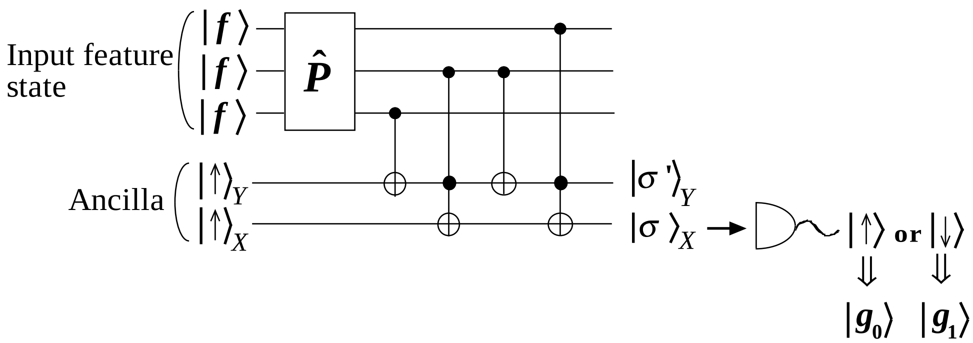

FIG. 2. A circuit realization of the optimal classifier in the case of three input samples. The three input samples are first transformed by $\hat{P}$, and then are processed interactively with two ancillary qubits via two controlled-NOT and two controlled-controlled-NOT gates. In our notation $\oplus$ is the operation which interchanges $|\uparrow\rangle$ and $|\downarrow\rangle$ and the filled circles are control lines, i.e. the operation $\oplus$ is applied iff all control lines are in state $|\downarrow\rangle$. Finally, the second ancillary qubit $|\sigma\rangle_X$ is measured in the basis $\{|\uparrow\rangle, |\downarrow\rangle\}$, and the two possible outcomes imply respectively that the best matched template is $|g_0\rangle$ or $|g_1\rangle$.

B. Template matching by state estimation

Another possible kind of semiclassical strategy based on the optimal state estimation of a qubit is also considered in Fig. 1. Quantum state estimation deals with how to evaluate unknown parameters of a quantum state as precisely as possible. This idea can be naturally applied to template matching: one can first perform a quantum state estimation to estimate the input feature state, and then compare this reconstructed state with the templates *classically*. Optimal state estimation of a qubit using N identically prepared states has been recently studied in Refs. [12–14]. In particular, discrete and finite element optimal POMs were found [13,14], and they maximize the following score

$$\bar{S}(N) \equiv \sum_m \frac{1}{2\pi}\int_0^{2\pi} d\phi\, \mathrm{Tr}\left(\hat{\Pi}_m \hat{f}(\phi)^{\otimes N}\right)|\langle f(\phi)|f(\phi_m)\rangle|^2. \tag{27}$$

where $|f(\phi_m)\rangle$ is a reconstructed state after the state estimation. The assignment of a guessed state $m \to |f(\phi_m)\rangle$ is now also to be optimized. This strategy was already described in Ref. [13], but we rephrase it here in a slightly different and more practical way according to the results of Ref. [16]. For convenience of calculation we introduce a new basis $\{|v_0\rangle, |v_1\rangle\}$ for which our great circle of feature states $|f(\phi)\rangle$ is the equator. We fix the basis vectors by requiring the templates to have the symmetrical form:

$$|g_0\rangle = \frac{1}{\sqrt{2}}\left(\mathrm{e}^{-i(\frac{\pi}{4}-\frac{\theta}{2})}|v_0\rangle + \mathrm{e}^{i(\frac{\pi}{4}-\frac{\theta}{2})}|v_1\rangle\right), \tag{28a}$$

$$|g_1\rangle = \frac{1}{\sqrt{2}}\left(\mathrm{e}^{i(\frac{\pi}{4}-\frac{\theta}{2})}|v_0\rangle + \mathrm{e}^{-i(\frac{\pi}{4}-\frac{\theta}{2})}|v_1\rangle\right). \tag{28b}$$

and the circle of feature states may be taken to be

$$|f(\phi)\rangle \equiv \frac{1}{\sqrt{2}}\left(\mathrm{e}^{-i\frac{\phi}{2}}|v_0\rangle + \mathrm{e}^{i\frac{\phi}{2}}|v_1\rangle\right). \tag{29}$$

where again, the parameter ϕ is uniformly distributed over $[0, 2\pi)$.

Let us also introduce M states equally spaced on the Bloch great circle (which will define our POM):

$$|f_m(\varphi)\rangle = \frac{1}{\sqrt{2}}\left[\mathrm{e}^{-i(\frac{\varphi}{2}+\frac{m\pi}{M})}|v_0\rangle + \mathrm{e}^{i(\frac{\varphi}{2}+\frac{m\pi}{M})}|v_1\rangle\right]; \quad (m = 0, 1, ..., M-1) \tag{30}$$

Here we have introduced a phase factor φ which determines the position of these symmetrical states relative to the fixed positions of the template states. The corresponding N-fold tensor product states are:

$$|F_m(\varphi)\rangle \equiv |f_m(\varphi)\rangle^{\otimes N} = \sum_{k=0}^{N}\sqrt{\frac{1}{2^N}\begin{pmatrix} N \\ k \end{pmatrix}}\mathrm{e}^{-i(N-2k)(\frac{\varphi}{2}+\frac{m\pi}{M})}|k\rangle_v, \tag{31}$$

where $\{|k\rangle_v\}$ is the symmetric bosonic basis for $\{|v_0\rangle, |v_1\rangle\}$. It can then be shown that the *square root measurement* $\{|\mu_m\rangle\langle\mu_m|\}$ based on the states $\{|F_m(\varphi)\rangle\}$, that is,

$$|\mu_m(\varphi)\rangle \equiv \left(\sum_{m=0}^{M-1}|F_m(\varphi)\rangle\langle F_m(\varphi)|\right)^{-\frac{1}{2}}|F_m(\varphi)\rangle, \tag{32}$$

and the associated guess $\{|f_m(\varphi)\rangle\}$, provides an optimal state estimation strategy when we take $M > N$ [16].

Thus, by applying the POM $\{|\mu_m(\varphi)\rangle\langle\mu_m(\varphi)|\}$, the input feature state is optimally reconstructed as one of the $|f_m(\varphi)\rangle$. Then one can *classically* compare this reconstructed state with the templates and pick up the template state which has the largest overlap with the reconstructed state. The above strategy for state *estimation* is optimal for any choice of φ. Indeed since our input state distribution is uniform, the score in eq. (27) will be independent of φ. But when we compare the reconstructed state with the templates, the resulting average score of template matching *will* depend on φ due to the fixed positions of the templates, i.e. different state estimation strategies which each give the same best possible score will generally give different scores for template matching via our classical method, and we should choose the best from our set of optimal estimation strategies. To complete our semi-classical template matching procedure we categorize the $|f_m(\varphi)\rangle$'s into two classes according to the template states, that is, condense the set $\{|\mu_m(\varphi)\rangle\langle\mu_m(\varphi)|\}$ into a two element POM $\{\hat{\Pi}_0^{\mathrm{EST}}(\varphi), \hat{\Pi}_1^{\mathrm{EST}}(\varphi)\}$, whose elements indicate that the best matched template is $|g_0\rangle$ or $|g_1\rangle$, respectively.

For simplicity let us assume that M is even. Then by symmetry, it is enough to consider φ in the range $[0, \frac{2\pi}{M})$. The categorization boundary is determined by the condition

$$|\langle g_0|f_m(\varphi)\rangle| = |\langle g_1|f_m(\varphi)\rangle| \tag{33}$$

The values of m with $|\langle g_0|f_m(\varphi)\rangle| \geq |\langle g_1|f_m(\varphi)\rangle|$ are categorized into the class of $|g_0\rangle$, and the others into that of $|g_1\rangle$. Noting that

$$|\langle g_0|f_m(\varphi)\rangle|^2 - |\langle g_1|f_m(\varphi)\rangle|^2 = 2\sin(\varphi + \frac{2m\pi}{M})\sin(\frac{\pi}{2} - \theta), \tag{34}$$

the binary categorization should then be

$$\hat{\Pi}_0^{\mathrm{EST}}(\varphi) = \sum_{m=0}^{\frac{M-2}{2}}|\mu_m(\varphi)\rangle\langle\mu_m(\varphi)|, \tag{35a}$$

$$\hat{\Pi}_1^{\mathrm{EST}}(\varphi) = \sum_{m=\frac{M}{2}}^{M-1}|\mu_m(\varphi)\rangle\langle\mu_m(\varphi)|. \tag{35b}$$

The average score for this strategy is

$$\bar{S}_{\rm EST}(N,M,\varphi) = \sum_{j=0}^{1}\frac{1}{2\pi}\int d\phi\, {\rm Tr}\left[\hat{\Pi}_j^{\rm EST}(\varphi)\hat{f}(\phi)^{\otimes N}\right]|\langle f(\phi)|g_j\rangle|^2 \tag{36a}$$

$$= {\rm Tr}\left[\hat{W}_0\hat{\Pi}_0^{\rm EST}(\varphi)\right] + {\rm Tr}\left[\hat{W}_1\hat{\Pi}_1^{\rm EST}(\varphi)\right], \tag{36b}$$

where

$$\hat{W}_0 = \frac{1}{2^{N+2}}\left[2\sum_{k=0}^{N}\begin{pmatrix} N \\ k \end{pmatrix}|k\rangle_{vv}\langle k| + \sum_{k=0}^{N-1}\begin{pmatrix} N \\ k \end{pmatrix}\sqrt{\frac{N-k}{k+1}}\left({\rm e}^{i(\frac{\pi}{2}-\theta)}|k+1\rangle_{vv}\langle k| + {\rm e}^{-i(\frac{\pi}{2}-\theta)}|k\rangle_{vv}\langle k+1|\right)\right] = \hat{W}_1^{\dagger}. \tag{37}$$

By a straightforward calculation we obtain

$$\bar{S}_{\rm EST}(N,M,\varphi) = \frac{1}{2} + \frac{\cos\theta\cos(\varphi-\frac{\pi}{M})}{2^N M\sin(\frac{\pi}{M})}\sum_{k=0}^{N-1}\begin{pmatrix} N \\ k \end{pmatrix}\sqrt{\frac{N-k}{k+1}} \tag{38a}$$

$$< \bar{S}_{\rm EST}(N,N+1,\frac{\pi}{M}) \tag{38b}$$

$$= \frac{1}{2} + \frac{\cos\theta}{2^N(N+1)\sin(\frac{\pi}{N+1})}\sum_{k=0}^{N-1}\begin{pmatrix} N \\ k \end{pmatrix}\sqrt{\frac{N-k}{k+1}}. \tag{38c}$$

The quantity $\bar{S}_{\rm EST}(N,N+1,\frac{\pi}{M})$ is compared with the optimal score $\bar{S}_{\rm OPT}(N)$ and the one $\bar{S}_{\rm MV}(N)$ obtained by the separable measurement plus majority voting scheme in Fig. 1 (for $\theta = 0$). As it can be seen, the strategy using the optimal state estimation followed by classical matching can be close to optimal for the region of small N, while as N increases it starts to deviate from the optimal one and becomes closer to the strategy of separable measurement plus majority voting.

The three strategies are schematically summarized in Figs. 3∼5. The quantum optimal strategy (Fig. 3) is realized by a collective measurement on the state $|f\rangle^{\otimes N}$ with binary outputs. This is made by dividing the state space $\mathcal{H}_B$ spanned by $|f\rangle^{\otimes N}$ into 2 parts according to the templates, and by successfully using entanglement effects in $\mathcal{H}_B$. On the other hand, the separable measurement plus majority voting scheme shown in Fig. 4 does not take any advantage of the entanglement which could be drawn from the state $|f\rangle^{\otimes N}$. In the optimal state estimation plus classical matching strategy shown in Fig. 5, the collective measurement first performed for estimating $|f\rangle$ also utilizes an entanglement effect. However, this is not the best way for *binary classification*. In fact, the optimal state estimation requires dividing the space $\mathcal{H}_B$ into at least $N+1$ parts. As N increases, one has to rely much more on the classical procedure to categorize the outputs into two classes. This is the reason why this strategy becomes ineffective for larger N. Intuitively any intermediate measurement prior to the final decision tends to degrade the total performance leading to a waste of input copies for a given average score level, so the process for the best binary classification should stay entirely in the quantum domain until the very final measurement.

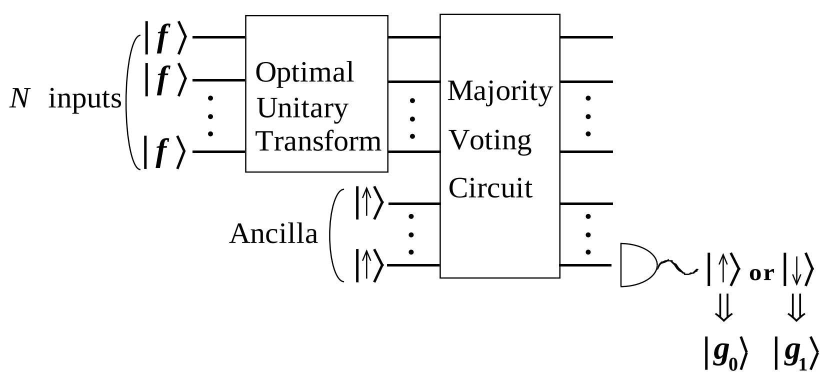

FIG. 3. A scheme of the optimal classifier for N input samples which is a generalization of Fig. 2. The binary classification of interest would eventually be turned into the measurement of a single qubit in the basis $\{|\uparrow\rangle, |\downarrow\rangle\}$.

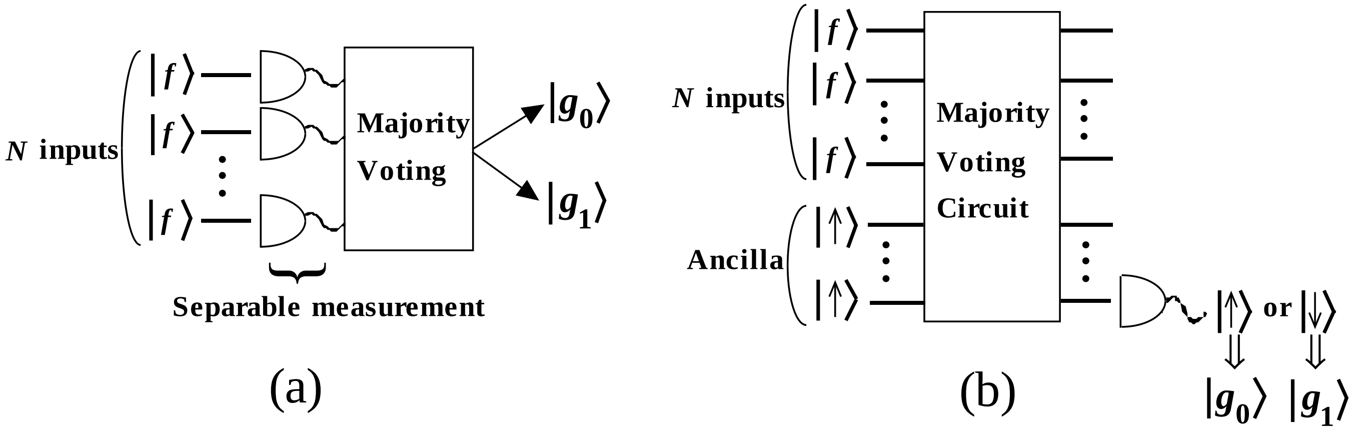

FIG. 4. Schemes of the strategy for the separable measurement plus majority voting. (a) is the direct translation of the POM which includes N measurements. But this can be translated into a measurement on a single ancillary qubit plus an additional circuit (majority voting circuit) beforehand as shown in (b). The majority voting circuit includes a series of C-NOT gates just as in Fig. 2.

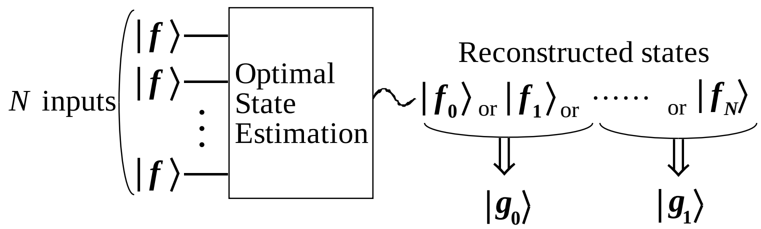

FIG. 5. A scheme of the strategy for the optimal state estimation plus classical matching. The optimal state estimation is a collective measurement on N identical copies of the sample state. By applying this, the input feature state is reconstructed. The output would be one of the $N+1$ candidates $|f_m\rangle$. Then one can compare this reconstructed state with the templates *classically*. This is actually a categorization of $|f_m\rangle$ into two classes according to the template states.

IV. MULTIPLE TEMPLATE MATCHING OF A TWO STATE SYSTEM

In the previous section we have assumed a single feature parameter ϕ and a minimum number (two) of templates. In this section we extend our model to allow for multiple template matching. Although binary template matching can be reduced to the diagonalization of the operator $\hat{W}_0 - \hat{W}_1$, there is no such straightforward method to find the optimal strategy in general cases. To keep the model tractable, we assume that the input feature state is a general qubit state depending now on two parameters,

$$|f\rangle = \mathrm{e}^{-i\frac{\phi}{2}}\cos\frac{\theta}{2}\,|\uparrow\rangle + \mathrm{e}^{i\frac{\phi}{2}}\sin\frac{\theta}{2}\,|\downarrow\rangle\,. \tag{39}$$

with a uniform *a priori* distribution over the whole Bloch sphere. Furthermore we suppose that only one of the parameters relates to the desired feature of $|f\rangle$, for example, the angle parameter ϕ around the $\hat{\sigma}_z$ axis, while the $\hat{\sigma}_z$ component itself is of no interest. The template states corresponding to this feature are assumed to be M states uniformly distributed around the great circle in the $x-y$ plane of the Bloch sphere, that is,

$$|g_m\rangle = \frac{1}{\sqrt{2}}\left(\mathrm{e}^{-i\frac{m\pi}{M}}\,|\uparrow\rangle + \mathrm{e}^{i\frac{m\pi}{M}}\,|\downarrow\rangle\right); \quad (m = 0, 1, ..., M-1). \tag{40}$$

As before we have N copies of the input state as

$$|F\rangle \equiv |f\rangle^{\otimes N} = \sum_{k=0}^{N}\sqrt{\begin{pmatrix} N \\ k \end{pmatrix}}\left(\mathrm{e}^{-i\frac{\phi}{2}}\cos\frac{\theta}{2}\right)^{N-k}\left(\mathrm{e}^{i\frac{\phi}{2}}\sin\frac{\theta}{2}\right)^{k}|k\rangle\,, \tag{41}$$

and generate the score operators based on this and the templates as

$$\hat{W}_m \equiv \frac{1}{4\pi}\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta |F\rangle\langle F| \, |\langle f|g_m\rangle|^2 \tag{42a}$$

$$= \frac{1}{2(N+1)}\left[\hat{I} + \sum_{k=0}^{N-1}\frac{\sqrt{(N-k)(k+1)}}{N+2}\left(\mathrm{e}^{-i\frac{m\pi}{M}}|k+1\rangle\langle k| + \mathrm{e}^{i\frac{m\pi}{M}}|k\rangle\langle k+1|\right)\right], \tag{42b}$$

We then seek the strategy to find the template which best matches with the given $|f\rangle$ in such a way to maximize the average score

$$\bar{S} = \sum_{m=0}^{M-1} \mathrm{Tr}\,(\hat{W}_m \hat{\Pi}_m). \tag{43}$$

As noted in section II this optimal template problem is equivalent to the problem of optimal discrimination of the set of mixed states $\frac{1}{2}\hat{W}_m$ taken with equal *a priori* probabilities $p_m = 1/M$.

The score operators evidently have the same symmetry as the templates, that is,

$$\hat{W}_m = \hat{V}^m \hat{W}_0 \hat{V}^{\dagger m}, \tag{44}$$

where

$$\hat{V} \equiv \sum_{k=0}^{N} \mathrm{e}^{-i\frac{(N-2k)\pi}{M}} |k\rangle\langle k|, \tag{45}$$

is a unitary representation of the group of integers modulo M on the $N+1$ dimensional bosonic subspace of N qubits. Indeed it is just the product representation $\hat{V} = \hat{v}^{\otimes N}$ where $\hat{v}$ is the operation of rotation of the one-qubit Bloch sphere by $2\pi/M$ about the z-axis. Now it is known [3] that for any group covariant set of states, the state discrimination problem always has an optimal strategy that is similarly group covariant, i.e. there will be an optimal POM of the form $\hat{\Pi}_m = \hat{V}^m \hat{\Pi}_0 \hat{V}^{\dagger m}$ and the optimality conditions reduce to [6]

$$\begin{aligned} &\text{(i')} \quad \hat{\Gamma} \equiv \sum_{m=0}^{M-1} \hat{V}^m \hat{W}_0 \hat{\Pi}_0 \hat{V}^{\dagger m} \text{ is hermitian,} \\ &\text{(ii')} \quad \hat{\Gamma} - \hat{W}_0 \geq 0. \end{aligned} \tag{46}$$

We have succeeded in deriving an optimal strategy only in the case that $M > N$, i.e. when the number of copies is less than the number of templates. This is again the square root measurement built from the templates $|G_m\rangle \equiv |g_m\rangle^{\otimes N}$, that is, the set $\{\hat{\Pi}_m = |\mu_m\rangle\langle\mu_m|\}$ with

$$|\mu_m\rangle \equiv \hat{G}^{-\frac{1}{2}}|G_m\rangle, \quad \hat{G} = \sum_{m=0}^{M-1}|G_m\rangle\langle G_m|. \tag{47}$$

In fact, by using the orthogonality relation

$$\sum_{m=0}^{M-1}\exp(i\frac{2m\pi}{M}n) = M\delta_{n,0} \qquad \text{for } -M < n < M, \tag{48}$$

(so $\hat{G}$ like $\hat{V}$ is diagonal in the $|k\rangle$ basis and $\hat{G}^{1/2}$ commutes with $\hat{V}$), we find that

$$|\mu_m\rangle = \hat{V}^m|\mu_0\rangle, \quad |\mu_0\rangle = \frac{1}{\sqrt{M}}\sum_{k=0}^{N}|k\rangle. \tag{49}$$

The optimality of this POM can then be proved by checking the conditions (i') and (ii') directly as follows. From Eq. (48), we obtain

$$\hat{\Gamma} = \frac{1}{2(N+1)}\left[\hat{I} + \sum_{k=0}^{N-1}\frac{\sqrt{(N-k)(k+1)}}{N+2}\left(|k\rangle\langle k| + |k+1\rangle\langle k+1|\right)\right], \tag{50}$$

and, consequently,

$$\hat{\Gamma} - \hat{W}_0 = \frac{1}{2(N+1)(N+2)} \sum_{k=0}^{N-1} \sqrt{(N-k)(k+1)} \Big[|k\rangle\langle k| + |k+1\rangle\langle k+1| - |k\rangle\langle k+1| - |k+1\rangle\langle k| \Big]. \tag{51}$$

Since each 2×2 matrix inside the brackets [...] in Eq. (51) has the eigenvalues 0 and 2 and is non-negative definite, so also is $\hat{\Gamma} - \hat{W}_0$ and (ii') of Eq. (46) holds. Condition (i') of Eq. (46) can be checked in a straightforward manner from Eqs. (49) and (51). The maximum average score does not depend on M and reads

$$\bar{S}_{\max}(N) = M \,\mathrm{Tr}\,(\hat{W}_0 \hat{\Pi}_0) = \frac{1}{2} + \sum_{k=0}^{N-1} \frac{\sqrt{(N-k)(k+1)}}{(N+1)(N+2)}. \tag{52}$$

We also note that $|G_0\rangle$ is the maximum-eigenvalue eigenstate of $\hat{W}_0$, i.e. the spectral decomposition is

$$\hat{W}_0 = \sum_{k=0}^{N} \omega_k \,|\omega_k\rangle\,\langle\omega_k|\,, \quad \omega_k = \frac{k+1}{(N+1)(N+2)}, \tag{53}$$

with $|G_0\rangle = |\omega_N\rangle$. This is especially interesting in view of the following theorem proved in [3]:
Theorem: Let G be a group and let $g \to \hat{V}_g$ be an *irreducible* representation of G on a d dimensional Hilbert space $\mathcal{H}$. Let $\{\hat{F}_g : g \in G\}$ be a collection of Hermitian operators on $\mathcal{H}$ such that $\hat{F}_g = \hat{V}_g \hat{F}_e \hat{V}_g^\dagger$ (where e is the identity of G). For any POM $X = \{\hat{X}_g : g \in G\}$ consider the function

$$Q(X) = \mathrm{Tr} \sum_g \hat{F}_g \hat{X}_g.$$

Let $Z = \frac{d}{|G|}\,|\phi\rangle\,\langle\phi|$ where $|\phi\rangle$ is the maximum eigenvalue eigenstate of $\hat{F}_e$ (and $|G|$ is the size of G).
Then Q is maximized by the POM $\{\hat{V}_g Z \hat{V}_g^\dagger : g \in G\}$.

Note that $\hat{V}_g\,|\phi\rangle$ is a maximum eigenvalue eigenstate of $\hat{F}_g$ so the theorem claims that the G covariant POM based on these $|G|$ eigen-directions is optimal. By irreducibility of the representation we have (via Schur's lemma) that $\sum_g \hat{V}_g \hat{A} \hat{V}_g^\dagger$ is a multiple of the identity for any operator $\hat{A}$. Thus the square root measurement construction does not alter these maximal eigen-directions when the representation is irreducible. In our template matching problem G is the group of integers modulo M and $\hat{F}_g$ correspond to the score operators $\hat{W}_m$. $\mathcal{H}$ is the $N+1$ dimensional bosonic subspace of N qubits and the group acts via $m \to \hat{V}^m$. This representation is *not* irreducible so the theorem does not apply. Yet we have shown that an optimal measurement is still obtainable from the maximum eigenvalue eigenstates of the score operators. In this case (of a reducible representation) the square root construction will give a non-trivial change in the directions of the maximal eigenstates, necessary to obtain a POM from them. This suggests a possible avenue of generalization for the above theorem of [3] which we will explore elsewhere.

In the other case $M \le N$, that is, when we can use a larger number of copies of the input than the number of templates, the optimal POM is more complicated. This should include elements with rank 2 or higher because of the requirement that $\sum_{m=0}^{M-1} \hat{\Pi}_m = \hat{I}$ in the $N+1$ dimensional bosonic subspace. We have not yet found a systematic way to construct such higher rank POMs. Here we discuss some simple cases.

The simplest case is $M = 2$, that is, binary classification. In this case, the two score operators commute and the strategy of separable measurement in the binary template basis on each copy plus majority voting turns out to be optimal (note that the binary template problem in section III had a different distribution of input states and the the two templates there were not required to be orthogonal).

The next simplest case is $M = N = 3$. The optimal POM is specified by

$$\hat{\Pi}_0 = \begin{pmatrix} \frac{1}{3} & a & c & 0 \\ a & \frac{1}{3} & b & c \\ c & b & \frac{1}{3} & a \\ 0 & c & a & \frac{1}{3} \end{pmatrix}, \tag{54}$$

with $a = (\sqrt{21} + \sqrt{5})/24$, $c = (\sqrt{35} - \sqrt{3})/24$, and $b = 6ac$, and the maximum average score is

$$\bar{S}_{\max}(N) = \mathrm{Tr}\,\hat{\Gamma} = \frac{5 + 3\sqrt{3}a + 3b}{10}. \tag{55}$$

This $\hat{\Pi}_0$ is derived by solving the equations for the condition (46)-(i') directly and then by picking up the solution satisfying the condition (46)-(ii'). $\hat{\Pi}_0$ is a rank 2 operator

$$\hat{\Pi}_0 = \lambda_+ \left|\lambda_+\right\rangle \left\langle\lambda_+\right| + \lambda_- \left|\lambda_-\right\rangle \left\langle\lambda_-\right| , \tag{56}$$

with $\lambda_+ = 0.964$ and $\lambda_- = 0.370$ and

$$\left|\lambda_+\right\rangle = 0.995 \left|\omega_3\right\rangle + 0.100 \left|\omega_1\right\rangle , \tag{57a}$$
$$\left|\lambda_-\right\rangle = 0.979 \left|\omega_2\right\rangle + 0.204 \left|\omega_0\right\rangle , \tag{57b}$$

where $\left|\omega_0\right\rangle, \left|\omega_1\right\rangle, \left|\omega_2\right\rangle, \left|\omega_3\right\rangle$ are the eigenstates of $\hat{W}_0$ corresponding to eigenvalues in increasing order (Eq. (53)). Thus, although the main component of $\hat{\Pi}_0$ comes from the maximum-eigenvalue eigenstate $\left|\omega_3\right\rangle$ of $\hat{W}_0$, the other eigenstates are also involved with appropriate weights.

Finally we mention the case of $M = 3$ and $N = 4$. The optimal POM is specified by

$$\hat{\Pi}_0 = \begin{pmatrix} \frac{1}{3} & a & c & 0 & -\sqrt{\frac{3}{8}}c \\ a & \frac{1}{3} & b & \sqrt{\frac{3}{8}}c & 0 \\ c & b & \frac{1}{3} & b & c \\ 0 & \sqrt{\frac{3}{8}}c & b & \frac{1}{3} & a \\ -\sqrt{\frac{3}{8}}c & 0 & c & a & \frac{1}{3} \end{pmatrix} , \tag{58}$$

with $b = \sqrt{29 + \sqrt{201}}/24$, $c = (\sqrt{67} - \sqrt{3})/(24\sqrt{2})$, and $a = 6bc$. The structure of $\hat{\Pi}_0$ is again of the form

$$\hat{\Pi}_0 = \lambda_+ \left|\lambda_+\right\rangle \left\langle\lambda_+\right| + \lambda_- \left|\lambda_-\right\rangle \left\langle\lambda_-\right| , \tag{59}$$

with $\lambda_+ = 1$ and $\lambda_- = 2/3$,

$$\left|\lambda_+\right\rangle = 0.992 \left|\omega_4\right\rangle + 0.115 \left|\omega_2\right\rangle + 0.044 \left|\omega_0\right\rangle , \tag{60a}$$
$$\left|\lambda_-\right\rangle = 0.984 \left|\omega_3\right\rangle + 0.178 \left|\omega_1\right\rangle , \tag{60b}$$

and the maximum average score is

$$\bar{S}_{\max}(N) = \frac{5 + 4a + 2\sqrt{6}b}{10} . \tag{61}$$

Generally speaking it is more difficult to find analytic solutions for the Bayes optimal strategy for mixed states, and one has to rely on numerical methods. The above examples indicate that the largest eigenvalue eigenstates of the score operator $\hat{W}_0$ should play an essential role in constructing the optimal POM (and maximizing the score), while smaller eigenvalue eigenstates can be regarded as perturbative correction terms. This might be helpful for considering efficient numerical algorithms for finding the optimal POM.

V. CONCLUDING REMARKS

We have considered the problem of quantum template matching, which is to find the template state that best matches a given input feature state. The quality of matching was taken to be the standard overlap of quantum states. This question was formulated in the context of quantum Bayesian inference and it was seen to be equivalent to the optimal discrimination of certain mixed states given in terms of score operators, each defined for a specific template state and including all the *a priori* information about the input.

In this paper, the simplest case of binary classification of a two state system with a single feature parameter was extensively studied. We constructed the optimal strategy in the $N + 1$ dimensional bosonic state space $\mathcal{H}_B$ spanned by the tensor products $\left|f\right\rangle^{\otimes N}$ of N identical copies of the input state. The optimal state estimation on $\left|f\right\rangle^{\otimes N}$ followed by a classical matching process does not provide the best strategy, and there is a different optimal use of entanglement for this particular binary classification problem. In the case of multiple template matching, the problem becomes more difficult and we derived the optimal strategies in a few illustrative cases.

As mentioned in the introduction, the procedure of conventional pattern matching consists of feature extraction, calculation of the discriminant function and classification. In the quantum context, however, it is not clear how to model such processes without an associated loss of useful information. For instance, to eliminate features of no concern one might simply project the input state onto the subspace spanned by the relevant states with the features of interest. But we saw in section III that a quantum measurement, that is, a projection of states, carried out before the final template decision, is generally detrimental to optimal performance. In this spirit, we dealt with the problem in the original Hilbert space without projecting the input states onto the subspace for the features of interest (section IV), and the whole process of quantum template matching was represented by a single POM. It is of course an open question to formulate quantum protocols in more physically comprehensive ways, e.g. involving a separate non-trivial feature enhancement process prior to classification, and to systematically derive optimal strategies for them.

But even without such additional infrastructure (e.g. feature enhancement) our problem of template matching has some interesting generalizations related to the role of classical versus quantum information in the formulation. In our formulation we have assumed that the input states (such as $|f\rangle$) are given as quantum information (i.e. unknown quantum states) whereas the template states ($|g_i\rangle$'s with known identities) are given as classical information. Furthermore our goal was to obtain the best template as classical information (i.e. knowledge of the identity of the best $|g_i\rangle$) via a suitable POM. The ingredients of this formulation can be relaxed in a variety of potentially interesting ways and here we mention two such ways:
(a) Instead of knowing the identities of the template states we may merely be given only some finite number (K) of copies of each template (so our original formulation is equivalent to $K = \infty$). One matching strategy would then be to apply state estimation to the sets of K copies and proceed as in our original formulation with the resulting estimated state identities. But this is unlikely to be optimal and we should consider a more fully quantum procedure which, for any input $|f\rangle$, identifies the best template class (still here as classical information) without attempting to obtain any further information about the identities of the template states themselves.
(b) A second more intrinsically quantum mechanical formulation of template matching involves obtaining the answer (i.e. the best matching template) only as quantum information. In this scenario we have a known prior distribution of inputs $\{|f_i\rangle\,; p_i\}$ and a known set of possible templates $\{|g_j\rangle\}$. Then given one (or more) copies of $|f_i\rangle$ we want to design a quantum process (i.e. a completely positive trace preserving map acting on the input) that outputs (one copy of) a quantum state σ_i of the form $|f_i\rangle \to \sigma_i = \sum_j p_{ij}\, |g_j\rangle\, \langle g_j|$ such that some suitable average score $\sum_{ij} p_i p_{ij} S(j|i)$ is maximized. Note that the formulation in our paper (of getting the best template as classical information) would provide one possible strategy since we can then construct the corresponding template state as a quantum state, but again, this would not be expected to be optimal since we produce a great deal of unwanted extra information in addition to the desired quantum output state.

There are yet further possible avenues for generalizing the formulation of the template matching problem. One is to study pattern classification with other kinds of matching criteria than fidelity, which would be chosen according to some specific application or purpose. For example, according to the quantum Sanov theorem (e.g. summarized in section IV of [17]) the quantum relative entropy $S(\hat{f}||\hat{g})$ between two quantum states provides an index for estimating the probability that the states will not be distinguished on the basis of an arbitrary measurement on N copies of the state. Thus the relative entropy provides an alternative, operationally intuitive, notion of "distance" between quantum states and we may consider maximizing the average relative entropy as our similarity criterion in template matching for some purposes.

The above remarks and generalizations show that the problem of template matching introduced in this paper is just the beginning of a fruitful area for further study. The formulation adopted in the paper is perhaps the simplest, in that it is closely related to an existing body of results on quantum Bayesian estimation. But a study of possible hybrid quantum-classical generalizations along the lines suggested above would provide a natural setting for characterizing new properties, and a deeper understanding, of quantum information itself, and especially the ways it fails to accord with familiar properties of classical information.

ACKNOWLEDGMENTS

The authors would like to thank T. Hattori for introducing them to the pattern matching problem and for giving valuable suggestions. They would also like to thank V. Buzek, A. S. Holevo, C. A. Fuchs, C. H. Bennett, O. Hirota, S. M. Barnett and A. Chefles for helpful discussions. A.C.'s research is supported by JISTEC under grant no. 199016. R.J. is supported by the U.K. Engineering and Physical Sciences Research Council.

[1] *Digital Pattern Recognition* (ed. by K. S. Fu, Springer-Verlag, New York, 1976).
[2] K. Fukunaga, *Statistical Pattern Recognition* (Academic Press, New York, 1990).
[3] A. S. Holevo, J. Multivar. Anal. **3**, 337 (1973).
[4] H. P. Yuen, R. S. Kennedy, and M. Lax, IEEE Trans. **IT-21**, 125 (1975).
[5] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[6] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
[7] A. Peres, *Quantum Theory: concepts and methods* (Kluwer Academic Publishers, Dortrecht, 1993).
[8] T. Cover and J. Thomas, *Elements of Information Theory* (John Wiley and Sons, New York, 1991).
[9] A. S. Holevo, Probl. Peredachi Inform. vol 9, no. 2, 31 (1973).
[10] A. S. Holevo, Theory Prob. Appl. vol. 23, 411 (1978).
[11] P. Hausladen and W. K. Wootters, J. Mod. Opt. **41**, 2385 (1994).
[12] S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995).
[13] R. Derka, V. Buzek, and A. K. Ekert, Phys. Rev. Lett. **80**, 1571 (1998).
[14] J. I. Latorre, P. Pascual, and R. Tarrach, Phys. Rev. Lett. **81**, 1351 (1998).
[15] R. F. Werner, Phys. Rev. A**58**, 1827 (1998).
[16] M. Sasaki, A. Carlini and A. Chefles, submitted to J. Phys. A (2000). Available at http://xxx.lanl.gov/abs/quant-ph/0011057
[17] V. Vedral and M. Plenio, Phys. Rev. A**57**, 1619 (1998).